AD-A102 167 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
CHARACTERIZATIONS AND GOODNESS OF FIT TESTS.(U)
JUN 81 F J O'REILLY, M A STEPHENS N00014-76-C-0475
UNCLASSIFIED TR-302 NL

END
DATE
FILMED
8 81
DTIC

6

CHARACTERIZATIONS AND GOODNESS OF FIT TESTS.

By

10 Federico J. O'Reilly [redacted] Michael A. Stephens

12 22

7 TECHNICAL REP[redacted]T NO. 302

11 2 JUN [redacted] 81

14

Prepared Under Contract
15 N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction. Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed random variables (i.i.d.r.v.) with distribution function F; we discuss the composite goodness-of-fit test given by the test of Ho: $F \in F_0$ versus the alternative Ha: $F \in F - F_0$, where F is some large class of distributions and F_0 is a parametric family to be tested. For example, F_0 may be the family of exponential distributions and F the family of continous distributions.

Many test procedures are based on characterizing properties of the family F_0, taking the form: a statistic $T(X_1, \ldots, X_n; F_0)$ has distribution Q, where Q is a unique distribution function, if and only if $F \in F_0$; the statistic T and the distribution Q may be univariate or multivariate. An example is a transformation from $X_1, \ldots, X_n$ to a statistic $T = (Z_1, \ldots Z_m)$, $(m < n)$, where the statistics Z_i are uniforms i.e. i i d with the uniform distribution between 0 and 1, which we shall write U(0,1), or where Z_i are ordere[d] uniforms, i.e. distributed like a random sample from U(0,1) which has then been placed in ascending order. Examples of these transformations are the Conditional Probability Integral Tranformat[ion] (CPIT) discussed in O'Reilly and Quesenberry (1973) and in Rincon Gallardo, Quesenberry and O'Reilly (1979) which give uniforms and the J and K transformations in Seshadri, Csörgö and Stephens (1969) which transform exponential random variables to a sample of ordered uniforms. For other examples of characterizations see Prohorov (19

and Seshadriand Csörgö (1969). In particular, the final step in the test of Ho is usually to calculate a test statistic T_1 based on the values of T, and compare T_1 with its distribution under Ho. The "if" part of the characterization above gives what is usually referred to as the distribution-free property of T. If the "only if" part fails, meaning that it might have distribution Q for some $F^* \varepsilon F - F_0$, it then follows that a test of Ho based on T will have power equal to the significance level α for detecting the alternative F^*. For a good test therefore, a characterization is needed.

Invariant Characterizations. Even when a characterization of F_0 forms the basis of a test, the values in T, and the value of T_1 calculated from T, may depend on the order in which the original x_i are indexed. When this is not the case we say that the characterization is invariant; then all statisticians following the test procedure will obtain the same value of the test statistic, and in general this has an intuitive appeal.

In this paper we concentrate on the problem of testing for the exponential distribution; F_0 is the family with members $F(x) = 1 - \exp(-x/\Theta)$, $x > 0$, with $\Theta > 0$ unknown. Two important characterization procedures studied by Seshadri Csörgö and Stephens (1969), called J and K transformations, are based on characterizations; J is not invariant, but K is, and these authors

show that K is superior in terms of power. Thus, more support is given to use invariant characterization procedures, at least for the exponential case.

It therefore seems worthwhile to develop a systematic approach which gives an invariant characterization. This is done in section 3 by means of the CPIT. The details are worked out for the exponential test, and it is shown that the subsequent characterization is connected with K while the usual CPIT is connected with J. We investigate these procedures, and a variant of K, by means of power studies. They extend and support the results of Seshadri, Csörgő and Stephens (1969). Thus a good approach to providing structure in goodness-of-fit procedures seems to be the search for invariant characterizations. Unfortunately, the general application of the invariant CPIT will be difficult computationally, and we conclude the paper with some comments on these problems.

2. Transformations involving characterization. Throughout the paper, we employ the following notation. The r.v. $X_1, X_2, \ldots, X_n$ will denote an unordered exponential sample whereas $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ will denote the corresponding sample, ordered in ascending order. Similarly $Z_1, Z_2, \ldots, Z_n$ will denote an unordered $U(0,1)$

sample and $Z_{(1)},\ldots,Z_{(n)}$ will stand for the corresponding ordered sample. When necessary, $X_1^*, X_2^*,\ldots,X_n^*$ and $Z_1^*, Z_2^*,\ldots,Z_n^*$ will also denote unordered exponential and $U(0,1)$ samples respectively.

The following transformations which are characterizations in a sense specified in each case, are well known and can be traced in several places in the literature. See for example, Sukhatme (1973), Galambos (1975), Ahsanullah (1978) and Stephens (1978). The J and K transfromations are those used by Seshadri , Csörgö and Stephens (1969). We give the transformations in symbolic form where the meaning is obvious.

The J transformation. This transforms n random exponentials into n-1 ordered uniforms.

$$(X_1,X_2,\ldots,X_n) \rightarrow \boxed{J} \rightarrow (Z_{(1)},Z_{(2)},\ldots,Z_{(n-1)})$$

with $$Z_{(j)} = \left(\sum_{i=j}^{j} X_i\right) \Big/ \left(\sum_{i=1}^{n} X_i\right), \quad j=1,\ldots n-1$$

The i.i.d. nonnegative r.v. $X_1,X_2,\ldots,X_n$ with positive mean are exponentially distributed if and only if $Z_{(1)},Z_{(2)},\ldots,Z_{(n-1)}$ are distributed as an ordered sample of size (n-1) from $U(0,1)$.

The N transformation.- (Sukhatme, 1937). This changes an ordered exponential sample into a unordered exponential sample with new values.

$$(X_{(1)}, X_{(2)}, \ldots, X_{(n)}) \rightarrow \boxed{N} \rightarrow (X_1^*, X_2^*, \ldots, X_n^*)$$

with $\qquad X^*_j = (n+1-j)\ (X_{(j)} - X_{(j-1)}) \quad , \quad j=1,\ldots,n \qquad (X_{(0)} \equiv 0)$

For $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, the unordered i.i.d. nonnegative r.v. $X_1, X_2, \ldots, X_n$ with positive mean are exponentially distributed if and only if $X_1^*, X_2^*, \ldots, X_n^*$ are i.i.d. r.v. exponentially distributed.

The E transformation.- This changes an ordered uniform sample into an unordered uniform sample with new values.

$$(Z_{(1)}, Z_{(2)}, \ldots Z_{(n)}) \rightarrow \boxed{E} \rightarrow (Z_1^*, Z_2^*, \ldots, Z_n^*)$$

with $\qquad Z_j^* = (Z_{(j)} / Z_{(j+1)})^j \quad , \quad j=1,\ldots,n \qquad (Z_{(n+1)} \equiv 1)$

For $Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)}$, the unordered i.i.d. r.v. $Z_1, Z_2, \ldots, Z_n$ are $U(0,1)$ if and only if $Z_1^*, Z_2^*, \ldots, Z_n^*$ are i.i.d. r.v. $U(0,1)$.

Another two transformations which are obvious characterizations will be needed.

The $\bar{1}$ transformation.- This takes a uniform sample into a new uniform sample by substracting each value from 1.

$$(Z_1, Z_2, \ldots, Z_n) \rightarrow \boxed{I} \rightarrow (Z_1^*, Z_2^*, \ldots, Z_n^*)$$

with $\quad Z_j^* = 1 - Z_j \quad , \quad j=1,\ldots,n.$

<u>The R transformation</u>.- This takes an exponential sample to a new sample by reversing the indexing of the original values. The new sample therefore has the same values as the old.

$$(X_1, X_2, \ldots, X_n) \rightarrow \boxed{R} \rightarrow (X_1^*, X_2^*, \ldots, X_n^*)$$

with $\quad X_j^* = X_{n-j+1} \quad , \quad j=1,\ldots,n$

<u>The K transformation</u>.- The K transformation is equivalent to first using N and then using J. Thus K, like J, transforms into n-1 ordered uniforms, and we write symbolically $K = J \circ N$.

By construction K is an invariant characterization whereas J is only a characterization. The power studies reported by Seshadri, Csörgö and Stephens (1969) for a wide variety of uniformity tests that follow the transformation, show the clear superiority of K. Moreover, it is also noted that J produces, for some alternatives, transformed values that are more evenly spread in the unit interval than if they were uniformly distributed; these are called <u>superuniforms</u>, and arise for example, when J maps samples from a half-normal distribution. J and K are connected with a uniforms-to-uniforms procedure first critically examined by Durbin (1961) and called G

by Seshadri, Csörgö and Stephens (1969). Both G and E have the property that they transform superuniforms into a set which is more clustered, and more likely to be detected by usual tests of uniformity.

<u>The CPIT transformation C</u>.- In O'Reilly and Quesenberry (1973) a transformation is proposed that produces (n-1) unordered uniforms $Z_1, Z_2, \ldots, Z_{n-1}$ from n unordered exponentials $X_1, X_2, \ldots, X_n$. The aim of such a transformation is to change the problem of testing exponentiality to that of testing uniformity. In that paper, the general CPIT procedure is given and is illustrated for several families. It is not shown that the procedure yields a characterization in general, and due to its construction, it is not an invariant procedure. For the exponential case, the CPIT transformation is given by

$$(X_1, X_2, \ldots, X_n) \rightarrow \boxed{C} \rightarrow (Z_1, Z_2, \ldots, Z_{n-1})$$

with $$Z_{j-1} = 1 - \left(1 - X_j / \sum_{i=1}^{j} X_i\right)^{j-1} , \quad j=2,\ldots,n$$

and it can easily be seen that this is the composition

$C = \bar{I} \circ E \circ J$. Hence for the exponential case C yields a characterization which is not invariant.

The Wang and Chang transformation W.- Recently, Wang and Chang (1977) have proposed the use of a characterization of the exponential distribution, by means of a transformation W to a uniform sample given by

$$(X_1, X_2, \ldots, X_n) \rightarrow \boxed{W} \rightarrow (Z_1, Z_2, \ldots, Z_{n-1})$$

with $$Z_j = \left\{ \left(\sum_{i=1}^{j} X_i \right) / \left(\sum_{i=1}^{j+1} X_i \right) \right\}^j \quad , \ j=1,\ldots,n-1$$

It is easily seen that $W = E \circ J$, and again is a characterization which is not invariant (Wang and Chang (1977), give an independent proof of the characterization).

3. An invariant CPIT. The idea behind the CPIT procedure is to apply the multivariate transformation due to Rosenblatt (1952) to the conditional distribution of the sample given the minimal sufficient statistic S for the family under consideration. In that way a set of independent uniforms is obtained. Since Rosenblatt's transformation requires absolute continuity of the multivariate distribution and since the conditional distribution of the whole sample given S is necessarily singular, one is forced to seek the maximum number of terms in the sample for which their conditional distribution given S is absolutely continuous (almost surely). In that way one gets the maximum number of uniforms.

If in this procedure, instead of considering the conditional distribution of $X_1, X_2, \ldots, X_n$ (the sample) given S, we actually consider the conditional distribution of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ given S, then the resulting transformation will be invariant. This procedure will be referred to as the invariant CPIT (ICPIT).

In what follows the ICPIT is worked out for the exponential case. Many of the attractive properties that are found in dealing with this case are due to the features of the exponential distribution. Nevertheless, the procedure outlined could prove useful in other cases as well, and other families are currently being studied.

In order to apply Rosenblatt's transformation to the conditional distribution of as many order statistics as possible while retaining absolute continuity, it is proposed that this maximum number should be n-1, just as in the ordinary CPIT.

Let S stand for $\sum_{i=1}^{n} X_i$, and also let $S^{(1)}$ stand for $\sum_{i=2}^{n} X_{(i)}$, $S^{(2)}$ for $\sum_{i=3}^{n} X_{(i)}$, etc. S is in this case the minimal sufficient statistic.

Suppose that the conditional distribution of $X_{(1)}, X_{(2)}, \ldots, X_{(n-}$ evaluated at $x_{(1)}, x_{(2)}, \ldots, x_{(n-1)}$ given S is absolutely continuous a.s., and denote it by $\tilde{F}n(x_{(1)}, x_{(2)}, \ldots, x_{(n-1)})$.

Applying Rosenblatt's transformation in this conditional setting, requires the computation of the marginal and conditional distributions, $\tilde{F}n(x_{(1)})$, $\tilde{F}n(x_{(2)}|x_{(1)}),\ldots,\tilde{F}n(x_{(n-1)}|x_{(1)},\ldots,x_{(n-2)})$ which are afterwards used to produce the independent U(0,1) transforms, $Z_1=\tilde{F}n(X_{(1)})$, $Z_2=\tilde{F}n(X_{(2)}|X_{(1)}),\ldots,$ $Z_{n-1}=\tilde{F}n(X_{(n-1)}|X_{(1)},\ldots,X_{(n-2)})$.

If all of the conditional distributions and the marginal distribution employed are absolutely continuous a.s., then the joint $\tilde{F}n(x_{(1)},\ldots,x_{(n-1)})$ is absolutely continuous a.s. and viceversa, so it will suffice to verify the absolute continuity a.s., of the (n-1) proposed distributions.

$\tilde{F}n(x_{(1)})$ may be computed directly or by means of Bayes' formula since the conditional distribution of $X_{(1)}$ given S can be expressed in terms of the conditional distribution of S given $X_{(1)}$, and the distribution of S, and these are well known. After doing the algebra we have

$$Fn\ (x_{(1)}) = \begin{cases} 0 & \text{if} \quad x_{(1)} < 0 \\ 1 - (1-nx_{(1)}/S)^{n-1} & \text{if} \quad x_{(1)} \in (0,S) \\ 1 & \text{if} \quad x_{(1)} > S \end{cases}$$

thus $Z_1 = 1 - (1 - nX_{(1)}/S)^{n-1}$. The next step is to obtain $\tilde{F}n(x_{(2)}|x_{(1)})$, which is $P[X_{(2)} < x_{(2)} | S, X_{(1)}]$ evaluated at $X_{(1)} = x_{(1)}$. Note that knowledge of S and $X_{(1)}$ is equivalent to knowledge of $S^{(1)}$ and $X_{(1)}$; thus $\tilde{F}n(x_{(2)}|x_{(1)})$ is $P[X_{(2)} < x_{(2)} | S^{(1)}, X_{(1)}]$ *evaluated at* $X_{(1)} = x_{(1)}$.

Given $X_{(1)}$, the other observations $X_{(2)}, X_{(3)}, \ldots, X_{(n)}$ are distributed as an ordered sample of size (n-1) from an exponential with origin $X_{(1)}$; thus one can obtain easily

$$P[X_{(2)} < x_{(2)} | S^{(1)}, X_{(1)}] =$$

$$= P[X_{(2)} - X_{(1)} < x_{(2)} - X_{(1)} \mid S^{(1)} - (n-1)X_{(1)}, X_{(1)}] \ ;$$

the above consideration yields

$$Fn(x_{(2)}|x_{(1)}) = \begin{cases} 0 & \text{if } \ x_{(2)} - x_{(1)} < 0 \\ 1 - [1 - (n-1)(x_{(2)} - x_{(1)})/\{S^{(1)} - (n-1)x_{(1)}\}]^{n-2} & \\ \text{if } \ x_{(2)} - x_{(1)} \in (0, S^{(1)} - (n-1)x_{(1)}) & \\ 1 \ \textit{elsewhere}; & \end{cases}$$

thus, $Z_2 = 1 - [1 - (n-1)(X_{(2)} - X_{(1)})/\{S^{(1)} - (n-1)X_{(1)}\}]^{n-2}$.

For the computation of $\tilde{F}n(x_{(3)}|x_{(1)},x_{(2)})$, the Markovian property of the order statistics and the fact that knowledge of $S, X_{(1)}$ and $X_{(2)}$ is equivalent to that of $S^{(2)}, X_{(1)}$ and $X_{(2)}$, yield the following result;

$$\tilde{F}n(x_{(3)}|x_{(1)},x_{(2)}) \text{ is } P[X_{(3)} < x_{(3)} \mid S, X_{(1)}, X_{(2)}]$$

evaluated at $X_{(1)}=x_{(1)}$, $X_{(2)}=x_{(2)}$. But $P[X_{(3)} < x_{(3)} | S, X_{(1)}, X_{(2)}]$ is equal to $P[X_{(3)} < x_{(3)} | S^{(2)}, X_{(1)}, X_{(2)}]$, and given $X_{(2)}$, $X_{(1)}$ is independent of $X_{(3)}$ and $S^{(2)}$, therefore

$$P[X_{(3)} < x_{(3)} | S^{(2)}, X_{(1)}, X_{(2)}] = P[X_{(3)} < x_{(3)} | S^{(2)}, X_{(2)}] \quad \text{a.s.}$$

Now, in order to compute this last conditional distribution we observe that given $X_{(2)}$, the set $X_{(3)}, X_{(4)}, \ldots, X_{(n)}$ is distributed as an ordered sample of size $(n-2)$ from an exponential distributio with origin $X_{(2)}$ so the previous approach is repeated.

By extending these results, we have the following

<u>Theorem</u> for $j=1,2,\ldots,n-1$, the r.v.

$$Z_j = 1 - \left\{ \frac{1-(n-j+1)X_{(j)}/(X_{(j)}+\ldots+X_{(n)})}{1-(n-j+1)X_{(j-1)}/(X_{(j)}+\ldots+X_{(n)})} \right\}^{n-j} \quad \text{where } X_{(0)} \equiv 0$$

are i.i.d. $U(0,1)$ if $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are ordered exponentials.

It can now be seen that the ICPIT can be written as the composition of previous characterizations: ICPIT = $\bar{I}$ o E o J o R o N.

<u>Comment</u>.- Recall that K = J o N, W = E o J and C = $\bar{I}$ o E o J, and define also M = J o R o N. If in subsequent tests for uniformity we use test statistics which give the same value for a set Z_i as for the corresponding set $1 - Z_i$, it is clear that C is equivalent to W = E o J (*i.e.* will give the same values of the test statistics) and ICPIT is equivalent to E o J o R o N. The test statistics used below, based on the empirical distribution function, have this property. Note also that since R simply indexes exponentials in reverse order, and does not give new values, me might expect K and M to have the same power properties. The step E which appears in the C composition will have the property that it takes superuniforms, which can sometimes be produced by J, into samples which would be declared significant using the usual (upper) tail of the test statistics. Thus we do not have to guard against the possibility of superuniforms as Seshadri, Csörgö and Stephens (1969) found necessary with J alone. On the other hand, E might sometimes take non-uniform samples into more evenly spaced observations, thus weakening the test, and we might find ICPIT, for some alternatives, giving a sample which is not so easily detected for non-uniforms as M, or its equivalent K.

4. Power Studies. These points have been examined by extensive power studies, using C, ICPIT and M. After the samples were taken from the alternatives listed, the statistics used for testing uniformity were the Cramér-Von Mises W^2, Watson U^2, Kolmogorov-Smirnov D, Kuiper V and Anderson-Darling A^2. Tables 1 and 2 give results one for sample size n=10 and the second for n=20, both with tests of size α=0.10. These should be compared for J and K given by Seshadri, Csörgö and Stephens (1969), and more extensive tables for K alone given by Stephens (1978).

Comments.- (a).- The presence of E after J in C does give upper tail significance for the samples from the half normal distribution where J alone gives superuniforms, as was conjectured above. C is still not as powerful in this case as J, using the lower tail of the test statistic, but it must be emphasized that one would not know that the lower tail is needed, so that C is to be preferred to J. (b) - As conjectured, M gives results very close to those given in Stephens (1978) for K for a wide range of alternatives. (c) - In general M (or K) give results a little better than ICPIT, which overall is better than C; in other words the tranformation related to ICPIT which gives ordered uniforms is preferred to ICPIT which gives unordered uniforms. (d).- K is therefore justified again as an effective chdracterization, and the ICPIT, which reproduced K , is

shown to provide a systematic approach to invariant characterizations. Unfortunately, for most families, the ICPIT will be difficult computationally; we hope that this work will stimulate further research into the general problem of finding invariant characterizations.

This work was partly supported by the National Science and Engineering Research Council of Canada, and partly by the U.S. Office of Naval Research, Contract No. N00014-76-C-0475.

Federico J. O'Reilly
IIMAS
Universidad Nacional Autonoma de Mexico
MEXICO

Professor M.A. Stephens
Mathematics Department
Simon Fraser University
Burnaby, BC
CANADA V5A 1S6

TABLE 1.- Simulated power in % for procedures C, ICPIT and M followed by W^2, U^2, K^2, D and A^2. Sample size n=10; significance level α=10; 1000 samples from each distribution.

	PROCEDURES														
	C = 1 E J					ICPIT =1 E J R N					M = J R N				
DISTRIBUTION	W^2	U^2	K^2	D	A^2	W^2	U^2	K^2	D	A^2	W^2	U^2	K^2	D	A^2
Exponential	10	9	10	11	10	11	9	10	10	10	10	10	11	10	10
U(0,1)	11	32	31	12	9	43	24	23	39	52	44	23	23	40	54
Chi Square 1 d.f.	22	27	26	24	42	37	21	22	34	41	42	25	24	39	53
Chi Square 3 d.f.	8	17	18	10	7	12	9	10	13	14	19	13	13	18	18
Chi Square 4 d.f.	10	29	28	12	7	18	12	13	17	21	35	18	18	31	34
Chi Square 6 d.f.	16	15	52	18	10	27	16	15	26	36	66	38	38	61	64
Log normal, σ=1	11	13	12	12	10	15	12	11	14	14	15	20	20	17	13
Log normal, σ=2	43	53	53	42	66	73	56	56	69	73	78	62	62	76	81
Log normal, σ=2.4	55	69	69	56	82	86	71	71	83	66	89	76	77	88	91
Beta, 1,4	10	12	13	10	8	8	10	10	8	10	12	11	10	12	12
Weibull .5	42	54	53	42	72	67	47	48	65	71	72	54	55	71	80
Weibull 2	15	54	52	17	10	34	19	18	30	44	70	37	36	63	69
Half normal	7	14	13	9	7	12	11	10	13	14	16	11	10	16	17
Half cauchy	26	27	27	26	33	47	37	38	46	47	48	43	42	47	48

TABLE 2.- Simulated power in % for procedures C, ICPIT and M followed by W^2, U^2, K, D and A^2. Sample size n=20; significance level α = .10; 500 samples from each distribution.

DISTRIBUTION	C = $\bar{I}$ E J					ICPIT = $\bar{I}$ E J R N					M = J R N				
	W^2	U^2	K	D	A^2	W^2	U^2	K	D	A^2	W^2	U^2	K	D	A^2
Exponential	11	11	10	11	11	9	9	10	9	9	8	9	9	8	9
U(0,1)	29	58	61	25	27	80	46	46	74	90	80	48	50	75	89
Chi Square 1 d.f.	38	48	45	38	63	52	33	31	48	56	62	38	40	57	75
Chi Square 4 d.f.	21	47	43	21	18	26	16	16	25	32	60	36	34	55	60
Chi Square 6 d.f.	50	85	82	48	50	39	21	21	35	52	94	67	68	90	93
Log normal, σ =1	16	20	19	16	15	21	19	18	21	22	24	31	30	22	22
Log normal, σ=2	71	82	81	68	87	91	84	83	89	91	95	87	87	93	96
Log normal, σ=2.4	86	94	94	85	97	97	92	93	96	97	98	95	96	98	99
Beta 1,4	10	14	13	10	9	15	11	11	14	16	15	10	10	15	16
Weibull .5	71	83	82	70	92	90	72	74	86	91	94	80	82	93	97
Weibull 2	56	89	85	51	56	55	29	27	49	70	99	52	52	97	100
Half normal	11	20	20	13	9	24	17	17	22	27	30	17	16	28	31
Half Cauchy	42	47	46	43	52	71	60	60	69	71	73	68	66	71	71

REFERENCES

Ahsanullah, M (1976) On characterization of the exponential distribution by order statistics. J. Appl. Prob. 13,818-822.

Csörgö M. and Seshadri, V. (1975) On the problem of replacing composite hipothesis by equivalent simple ones. Rev. ISI 38.

Durbin J. (1961) Some methods of constructing exact tests Biometrika, 48, 41-55.

Galambos J. (1975) Characterizations of probability distributions by properties of order statistics. Statist. Dist. in Scient. work, (G.P. Patil et. al, eds); Reidel; Dordrecht, 3, 71-102.

O'Reilly F.J. and Quesenberry C.P. (1973) The conditional probability integral transformation and applications to obtain composite chi square goodness-of-fit tests. Ann. Statist. 1,74-83.

Prohorov, Y, V (1966) Some characterization problems in statistics. Proceed. of 5th Berkeley Symp. on Math. Stat. and Prob. 1,341-349. Univ. Cal. Press; Berkeley, California.

Rincón Gallardo, S, Quesenberry C.P. and O'Reilly F.J. (1979) Conditional probability integral transformations and goodness-of-fit tests for multivariate normal distributions. Ann. Statist. 7,1052-1057.

Rosenblatt M. (1952) Remarks on a multivariate transformation. Ann. Math. Statist. **23**, 470-472.

Seshadri V, Csörgö M. and Stephens M.A. (1969) Tests for the exponential distribution using Kolmogorov-type statistics. J. Roy Stat. Soc. Ser B. **31**, 499-509.

Sing N.J. and Oliker V.I. (1977) On minimum variance estimation and characterization of probability Tech. Research Report No. 44. Dept. Statistics, Temple University, Penn.

Stephens M.A. (1978) Goodness-of-fit tests with special reference to tests for exponentiality. Tech. Report No. 262. Dept. of Statistics, Stanford University, Ca.

Sukhatme P.V. (1937) Tests of significance for samples of the X^2 population with two degrees of freedom. Ann. Eugenics, **8**, 52-56.

Wang Y.H. and Chang S.A. (1977) A new approach to non-parametric tests of exponential distribution with unknown parameters. Th. App. Reliability **2**, 235-258.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 302	2. GOVT ACCESSION NO. AD-A102167	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CHARACTERIZATIONS AND GOODNESS OF FIT TESTS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) FEDERICO J. O'REILLY & MICHAEL A. STEPHENS		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office Of Naval Research Statistics & Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE JUNE 2, 1981
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Characterizations; exponential distribution; goodness of fit tests; invariant characterizations and goodness of fit.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TR # 302

CHARACTERIZATIONS AND GOODNESS OF FIT TESTS

In this article a systematic approach to providing goodness of fit tests is discussed, for the composite goodness of fit problem of testing that the distribution F of a random sample comes from a parametric family F_0. Characterization procedures are emphasized, and it is shown that, at least for the exponential case, invariant characterizations appear to be better than those which are not invariant. A general technique is developed for producing invariant characterizations and for the exponential case it is shown how these are related to characterizations already in the literature. Power studies are given to examine the tests based on both invariant and non-invariant characterizations.

DATE

ILME